The repetitive landscape of the Robusta coffee genome

Romain Guyot [1,2], Chaparro Christian [3], Stéphanie Bocs [5], Alexis Dereeper [4], Véronique Jamilloux [6], Douglas Silva Domingues [2], Philippe Lashermes [4].

[1] *Institut de Recherche pour le Développement (IRD), UMR DIADE (CIRAD, IRD, UM2), BP 64501, 34394 Montpellier Cedex 5, France;* [2] *Instituto Agronômico do Paraná, Coordenação de Pesquisa, Laboratório de Biotecnologia Vegetal, 86047-902, Londrina, PR, Brasil;* [3] *2EI UMR5244 Université de Perpignan Via Domitia, UMR 5244 CNRS Ecologie et Evolution des Interactions (2EI), 52 Avenue Paul Alduy, 66860 Perpignan Cedex, France;* [4] *IRD UMR RPB, DIVA, BP 64501, 34394 Montpellier Cedex 5, France;* [5]*CIRAD (Centre de coopération internationale en recherche agronomique pour le développement, UMR AGAP, F-34398 Montpellier, France;* [6] *INRA, Unité de Recherches en Génomique-Info (UR INRA 1164), Centre de recherche de Versailles, 78026 Versailles Cedex, France*

Email: *romain.guyot@ird.fr*

The Robusta Coffee (*Coffea canephora*) is the most important worldwide commodity produced by tropical countries. Its diploïde genome sequence (2x=2n=22) is now available (http://coffee-genome.org). Transposable elements (TEs) are the most repetitive sequences in plant genomes, while their activity has a wide impact on genome composition, structure and evolution. The current challenge is to identify and characterize the repeated content of Robusta and to compare them to sequenced crop species.

Here we present the identification and classification of TEs in the 568 Mb of genomic sequences released by the Coffee Genome Sequencing Consortium, using a combination of *ab initio,* similarity and structure search approaches. We discovered that about half of the genomic sequences produced are composed of TEs, similarly to other sequenced crop species such as banana, papaya, castor bean and soybean. Class I LTR retrotransposons represent the vast majority of identified elements, accounting for 42% of the genome assembly, while transposons represent only 3.8%. Interestingly active non-autonomous LTR retrotransposons elements were detected and classified into a new subgroup of non–autonomous elements containing a *capsid* domain but lacking the region involved in their mobility. This newly discovered class of transposable elements called TR-GAG was found conserved in most of plant species investigated. Finally in an attempt to study conservation of LTR retrotransposons between coffee and Angiosperm genomes we identified outstanding conservations of several families across distantly related plant species, suggesting that conservation of such elements or horizontal transfer events might be more frequent than currently recognized.

.